Data Deposit Workflow

Data under 200 MBs can be stored within the IUSW repository. If your data is in Excel spreadsheets, we recommend depositing both .xlsx and .csv versions.

Data over 200 MBs is stored in IU's Scholarly Data Archive (link), with download links mapped to new IUSW records.

IU researchers can transmit any size data to the IUScholarWorks team by
- dropping files or zipped packages into a Box widget embedded in our wiki: https://wiki.dlib.indiana.edu/x/swRNHw
- moving files (or PUT) to our public SDA directory at **/hpss/i/u/iuswdata/dropbox/**

  (before moving files, please use command "umask 000" for the session to ensure we'll have permission to move the files)

Ideally, data will be bundled as a single archive file (.zip, .tar, .tgz, etc.)

Data stored in SDA requires additional documentation
- A *manifest* file that contains a checksum for the data
- A *README* that includes metadata and other critical information for the data

Both of these files should be uploaded as regular files (bitstreams) of the record in IUScholarWorks.

Manifest file

Any IUSW item with material stored in the Scholarly Data Archive should have a corresponding manifest file uploaded as a document (bitstream). The file contents should look like the following:

```
---------------

#manifest file for LEADII Vortex2 Archive Dataset
(http://hdl.handle.net/2022/uniqueid)

LEADII-Vortex2-dataset.tar.gzaa 30000000000 f2b425d71600959318e214450b7321f4
(this # is a checksum)

LEADII-Vortex2-dataset.tar.gzab 30000000000
cc90092d9fdffa6466c6a8bfc3b8ce63 (this # is a checksum)

---------------
```

The IUScholarWorks staff can run the checksum and create a manifest file for any data transmitted to us. The checksum can be created using the command line of Terminal (Apple) $ md5 /path/to/data_package or Linux $ md5sum /path/to/data_package.

README file

The README *file should also uploaded as a regular bitstream into the DSpace record*, and it should always be named "README.txt" [link]. The file contents should look similar to the following:

```
README FILE Name

Created by:
Name
Address
Address

URI [IUScholarWorks Handle, e.g., http://hdl.handle.net/2022/20995]

DOI [in URL format, e.g., https://doi.org/10.5967/K8SF2T3M]

---------------------------------------

FILE LIST

Vortex2-dataset.tar.gzaa

Vortex2-dataset.tar.gzab

[etc]

---------------------------------------

FILE INFORMATION

[List any special considerations for software/websites used to create the
files, file formats, etc. For example, "These tar files were produced by a
///////////////// and can be viewed with ////////////// software packages."]

---------------------------------------

RESEARCH/ QUESTION(S) and/or METHODOLOGY

[text]

---------------------------------------

DATA-SPECIFIC INFORMATION

[Variable names or abbreviations that aren't self-evident, units of
measurement, definitions for codes or symbols, etc.]

---------------------------------------
```

```
CHANGELOG

[each time you take a new step that will eventually change your output files
from the source text, document it here.]

0. Cleaned up data

1. Uploaded files

2. Named stopwords list using Options menu

3. etc

---------------------------------------

COPYRIGHT & LICENSING INFORMATION

Source data was created by [data source name & contact information here].

This data is licensed for reuse under a Creative Commons Attribution 4.0
license [or other relevant data license--remember that data cannot be
copyrighted in most cases].

---------------------------------------

LIMITATIONS

[note any limitations to your metholodology, problems encountered with the
software, etc.]
```

Add appropriate PURLs to DSpace items

The PURL is generated automagically based on the file name and location. For example a file named upload-808.zip in a directory structure 2022/123456 would be accesible at http://purl.dlib.indiana.edu/iusw/data/2022/123456/upload-808.zip. In other words, append the handle/filename to this root URL: **http://purl.dlib.indiana.edu/iusw/data/2022/**.The DC field to add these PURLs is **dc.relation.uri.**